Stat. Meth. & Appl. (2007) 16:39–49
DOI 10.1007/s10260-006-0018-4

ORIGINAL ARTICLE

A Bayesian approach to model interdependent event histories by graphical models

Emanuela Dreassi · Anna Gottard

Accepted: 17 March 2006 / Published online: 13 June 2006

Abstract In event history analysis, the problem of modeling two interdependent processes is still not completely solved. In a frequentist framework, there are two most general approaches: the *causal approach* and the *system approach*. The recent growing interest in Bayesian statistics suggests some interesting works on survival models and event history analysis in a Bayesian perspective. In this work we present a possible solution for the analysis of dynamic interdependence by a Bayesian perspective in a graphical duration model framework, using marked point processes. Main results from the Bayesian approach and the comparison with the frequentist one are illustrated on a real example: the analysis of the dynamic relationship between fertility and female employment.

Keywords Event history analysis · Interdependent marked point processes · Frailty models · Semi-parametric Bayesian models

1 Introduction

Event history analysis deals with random sequence of events and the dependence of time to an event on some explanatory variables. An appealing approach to treat event history data is due to Arjas (1989), that considers them as the sample path of a marked point process (MPP). Sometimes, complex systems can arise when two or more different MPPs can interact with one other. In the frequentist literature, much work has been developed focusing on

E. Dreassi (✉) · A. Gottard
Department of Statistics "G. Parenti", University of Florence,
Viale Morgagni 59, 50134 Florence, Italy
e-mail: dreassi@ds.unifi.it

A. Gottard
e-mail: gottard@ds.unifi.it

interdependence between survival times. For the general framework of event history analysis, there are two main approaches: the *causal approach* by Blossfeld et al. (1995) and the *system approach*, utilized by Tuma and Hannan (1984), but firstly described by Cox and Lewis (1972). By these approaches, compared in Moro and Gottard (1999), it is possible to analyze the mutual dependence of two event histories, influenced by a set of explanatory variables (time-constant) and processes (time-varying).

As known (see for example Cox and Wermuth 1996; Lauritzen 1996), the chain graphical models provide an useful tool to investigate the existence of both symmetric and asymmetric relationships among random variables and the standard theory can be extended to a class of MPPs. The conditions under which chain graphs could be able to identify the set of conditional independence relationships for the class of models including both MPPs and random variables are explored in Gottard (2006). The class of models for event history data representable by chain graphs are called *graphical duration models*. Briefly, we remind that a graphical duration model is a block recursive model where a MPP is represented as a single node in a graph ($\overleftarrow{\circ}$). Each block of such a graph can contain only processes or variables; blocks containing variables are always before blocks with processes. Cycles are forbidden. The way of reading conditional independencies from the graphs is given by a version of the block-recursive Markov property. The subgraph containing only variables is treated as usual, according to LWF Markov properties (see Frydenberg 1990; Lauritzen and Wermuth 1989). We remand to Gottard (2006) for a more extensive presentation of graphical duration models.

The aim of this paper is to evaluate a possible further development of graphical duration models within a Bayesian approach, by modeling the interdependence between two event histories by way of a semi-parametric hierarchical Bayesian model. In fact, the recent growing interest in Bayesian statistics suggests some interesting works on survival models and event history analysis in a Bayesian perspective (see for example Ibrahim et al. 2001). We extend the semi-parametric Bayesian intensity model for a single MPP proposed by Härkänen et al. (2000) to the case of two interdependent MPPs. The resulting model is a semi-parametric Cox model for two MPPs, whose conditional independence structure can be described by a chain graph. The parametric part of the model proposed includes both observed and unobserved explanatory variables to take into account individual frailty. The interdependence is considered in similar way of the causal approach, making the hypothesis that the two processes depend on each other by means of their past histories.

The paper is organized as the follows. In Sect. 2 we define notation and recall some useful definitions. In Sect. 3 the semi-parametric Bayesian intensity model for two interdependent MPPs and its estimation method are discussed. Section 4 applies the model to real data from INF-2 survey about control and expectations of fertility in Italy, to analyze relationship between fertility and female employment. In this section, moreover, we compare results obtained by the frequentist system approach (Gottard 2006) and Bayesian approach. Discussion and conclusions are reported in Sect. 5.

2 Some preliminaries

We define notation and we recall some definitions about the class of MPPs of our interest (for further details see Arjas 1989) and for conditional independence involving one or two MPPs (for further details see Aalen 1987). Briefly, a MPP $Y(t,m)$ is as a countable set of pairs $\{T_n, M_n\}_{n\in\{1,2,\ldots\}}$, where $T_n \in [0,\tau]$ is the occurrence time of the n-th event and the mark $M_n \in \mathcal{M}$ is an additional variable associated with each T_n describing the event type; the mark space is usually rather small. For instance, let $Y(t,m)$ be the process describing the individual occupational status while its mark space is $\mathcal{M} = \{m_1, m_2\} =$ {to find a job, to leave a job}. From now on, we assume that the unknown probability distribution of the MPP $Y(t,m)$ is absolutely continuous and that it can be conveniently specified by its *mark-specific hazard functions* $h_m(t)$, $m \in \mathcal{M}$. Next, we define H_t as the *past history* of the MPP, constructed in such a way that H_{t^-} contains the relevant history of $Y(t,m)$ up to t. We call *conditional mark-specific hazard function* $h_m\,(t \mid H_{t^-})$. Heuristically, $h_m\,(t \mid H_{t^-})\,dt$ can be viewed as the conditional probability of a type m event in $[t, t+dt)$, given the pre-t history H_{t^-}.

Now, we can say that a MPP $Y(t,m)$ is independent of a random variable Z if its probability distribution is not a function of Z. Therefore, given that the probability distribution of a MPP can be entirely written in terms of its marks specific hazard functions, using Dawid's notation (see Dawid 1979), we can say that

$$Y(t,m) \perp\!\!\!\perp Z_2 \mid Z_1 \text{ iff } h_m\,(t \mid Z_1, Z_2) = h_m\,(t \mid Z_1) \tag{1}$$

for all $t > 0$ and $m \in \mathcal{M}$. The definition of independence between two MPPs, as given by Schweder (see Schweder 1970), for Markov processes, and extended by Aalen (see Aalen 1987) for a more general case, is more complex and takes into account two different kinds of relationship: the *local independence* and the *stochastic independence*. To explain such definition, let $X(t,m_X)$ and $Y(t,m_Y)$ be two MPPs allowing for Doob–Meyer decomposition; let $H^X_{t^-}$ and $H^Y_{t^-}$ be their pre-t histories. Then, $X(t,m_X)$ is locally independent of $Y(t,m_Y)$, if its probability distribution is not a function of $H^Y_{t^-}$. Therefore, the definition of local independence, since it involves the probability of only one MPP, is not symmetric. Moreover, two processes $X(t,m_X)$ and $Y(t,m_Y)$ are *mutually locally independent* if and only if they are also stochastically independent. To distinguish stochastic from local independence we use the symbol $\perp\!\!\!\perp_{\rightarrow}$ for the latter one, so that $X(t,m_X) \perp\!\!\!\perp_{\rightarrow} Y(t,m_Y)$ means that $X(t,m_X)$ is locally independent of $Y(t,m_Y)$ as proposed in Gottard (2006). Similarly, the symbol $\not\perp\!\!\!\perp_{\rightarrow}$ states for local dependence. By the definition given, one can deduce that, for local independence

$$X(t, m_X) \perp\!\!\!\perp_{\rightarrow} Y(t, m_Y) \quad \text{iff } h_{m_X}\left(t \mid H^X_{t^-}, H^Y_{t^-}\right) = h_{m_X}\left(t \mid H^X_{t^-}\right) \tag{2}$$

for all $t > 0$ and $m_X \in \mathcal{M}_X$. Similarly,

$$X(t, m_X) \perp\!\!\!\perp Y(t, m_Y) \quad \text{iff } \begin{array}{c} h_{m_X}\left(t \mid H^X_{t^-}, H^Y_{t^-}\right) = h_{m_X}\left(t \mid H^X_{t^-}\right) \\ \text{and} \\ h_{m_Y}\left(t \mid H^Y_{t^-}, H^X_{t^-}\right) = h_{m_Y}\left(t \mid H^Y_{t^-}\right) \end{array} \tag{3}$$

for all $t > 0$, $m_X \in \mathcal{M}_X$ and $m_Y \in \mathcal{M}_Y$. Conditional independence and conditional local independence, given a set of random variables $\mathbf{Z}$, definitions can be directly derived from (1), (2) and (3).

3 The model

The probabilistic model assumed for each MPP taken into account can be considered a version of a semi-parametric Cox model.

Let us consider two MPPs $X(t, m_X)$ and $Y(t, m_Y)$, with $t > 0$, $m_X \in \mathcal{M}_X$ and $m_Y \in \mathcal{M}_Y$. Their mark-specific hazard functions $h_{m_X}(t)$ and $h_{m_Y}(t)$ are both decomposed as product of a non-parametric part shared by the subjects, the baseline hazard rate ($h_0^{m_X}(t)$ and $h_0^{m_Y}(t)$, respectively), and a parametric part functions of individual factors. To model the interdependence among the two MPPs, according with definitions given above (2–3), we assume that in each instant of time t, the two MPPs depend one on another by their past history.

In the following we define the model considering the MPP $Y(t, m_Y)$; similar definition could be arranged for the other MPP.

The mark-specific hazard function $h_{m_Y}\left(t \mid Z, H^Y_{t^-}, H^X_t\right)$ are equal to

$$h_0^{m_Y}(t) \;\; \exp\left\{\beta'_{m_Y}\mathbf{Z} + \gamma^Y_{m_Y} \; H^Y_{t^-} + \gamma^X_{m_Y} \; H^X_t\right\} \;\; U_{m_Y} \tag{4}$$

where $\mathbf{Z}$ are the covariates, H^Y_t and H^X_t the past histories of the two MPPs and U_{m_Y} is a latent variable taking into account individual frailty specific for mark m_Y.

According with (2), $Y(t, m_Y) \perp\!\!\!\perp_{\rightarrow} X(t, m_X)$ if $\gamma^X_{m_Y} = 0$ for all $m_Y \in \mathcal{M}_Y$ and $t > 0$, while $X(t, m_X) \perp\!\!\!\perp Y(t, m_Y)$ if both $\gamma^Y_{m_X} = 0$ and $\gamma^X_{m_Y} = 0$ for all $m_Y \in \mathcal{M}_Y$, $m_X \in \mathcal{M}_X$ and $t > 0$.

Following Arjas and Gasbarra (see Arjas and Gasbarra 1994), we assume that baseline hazard functions $h_0^{m_Y}(t)$ in (4) are piecewise constant with a support $[0, T_{\max}]$ (an approximation facilitating the numerical integration of the posterior distribution)

$$h_0^{m_Y}(t) = \sum_k a_k \mathbb{1}_{t \in (T_k, T_{k+1}] \cap (0, T_{\max}]}$$

where $a_k \geq 0$ are the levels and T_k the jumps points on $(0, T_{\max}]$ for $Y(t, m_Y)$.

For each mark of each MPP, we define as prior distribution of the jump points $T_1 < \cdots < T_k < \cdots < T_{\max}$ a Poisson process on $(0, T_{\max}]$ with hyperparameter μ. The prior distribution of the levels $a_1, \ldots, a_k, \ldots$ is specified inductively as follow: we suppose that the initial level $a_1 \sim \Gamma(\alpha_0, \beta_0)$, and that $a_k \mid a_1, \ldots, a_{k-1} \sim \Gamma(\alpha, \alpha/a_{k-1})$.

The hyperparameters α and μ control the a priori fluctuation of the function $h_0^{m_Y}(t)$ and α_0 and β_0 the expectation and the variance of the initial level. The hyperparameters $\alpha_0, \beta_0, \alpha, \mu$ assume values reflecting a vague prior knowledge for $h_0^{m_Y}(t)$; for each mark the hyperparameter values are $\alpha_0 = 0.01$, $\beta_0 = 0.1$, $\alpha = 0.1$ and $\mu = 0.1$, so that prior mean for a_1 is $\alpha_0/\beta_0 = 0.1$, the coefficient of variation is $1/\sqrt{\alpha_0} = 1/\sqrt{0.01}$ and the conditional standard deviation is $a_{k-1}/\sqrt{0.1}$. This construction of the baseline hazard functions $h_0^{m_Y}(t)$, given that the jump points are not a priori fixed, is equivalent to a non parametric definition.

In the parametric part of the mark specific hazard functions (4), the regression parameters β_{m_Y}, $\gamma_{m_Y}^{Y}$ and $\gamma_{m_Y}^{X}$ coefficients are assumed a priori to be Normal distributed centered on zero with high variance.

To take into account unobserved heterogeneity, the frailty term U_{m_Y} is introduced (on a multiplicative way) in the parametric part of each mark-specific hazard function. The frailty terms are a priori assumed to be conditionally independent and identically distributed, given a hyperparameter ϕ; $U_{m_Y} \sim \text{Gamma}(\phi, \phi)$, so the a priori expectation and variance are respectively one and ϕ^{-1}. To control the variance it is assumed that parameter ϕ is a random variable with prior distribution Gamma(2,2). The frailty terms U are assumed to be independent of the other explanatory variables $\mathbf{Z}$.

The model parameters are estimated by using Markov chain Monte Carlo (MCMC) integration techniques as Metropolis–Hastings algorithm, implemented using BITE (see Härkänen 2002). The algorithm seems to converge after 5,000 iterations, however, given also the very high number of (non monitored) parameters in the model, we decided to discard the first 100,000 iterations (burn-in) and to store for estimation 5,000 samples.

4 A real example

In this section, a Bayesian graphical duration model is applied to study the relationship between fertility and female employment. The data used are from Inf-2 survey about control and expectations of fertility in Italy (De Sandre et al. 1997): this is a retrospective study carried out in the period November 1995 – January 1996 on a sample of 4,533 women aged 20–49 years. The data set collect information on partnerships, fertility, employment and socio-demographic characteristics. In this example, it seems interesting to investigate the relationship between fertility and female labor force participation viewed as two interdependent marked point processes, possibly influenced by some original family characteristics.

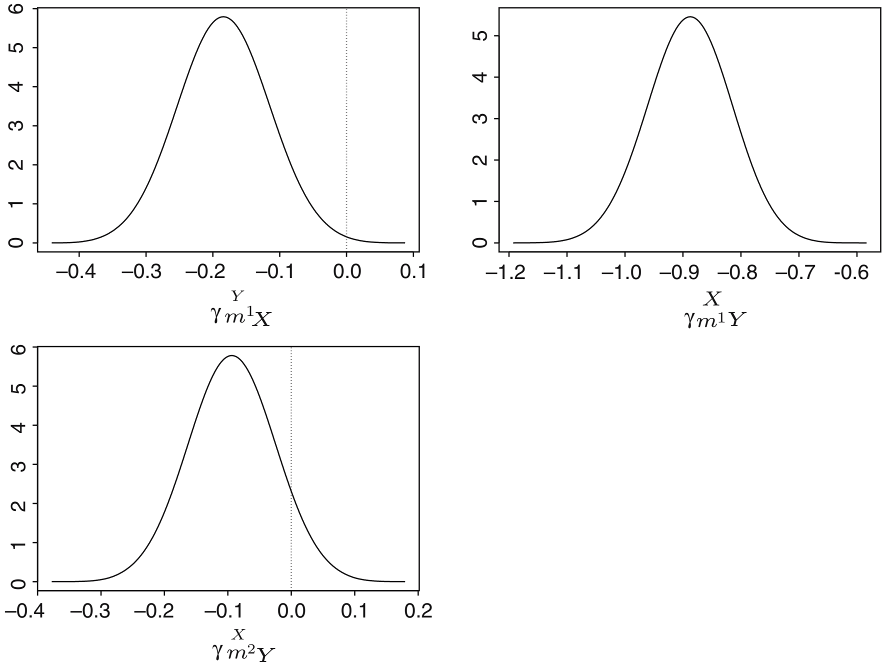

Fig. 1 Posterior distributions of parameters γ representing interdependence

Hence, let $X(t, m_X)$ denote fertility MPP, with only one mark representing the event is $m1_X =$"to have a child". Let $Y(t, m_Y)$ be the MPP describing labor-force participation, its mark space $\mathcal{M}_Y = \{m1_Y, m2_Y\}$, where $m1_Y$ indicates the event "to find a job" and $m2_Y$ "to leave a job"; so the process $Y(t, m_Y)$ has two marks.

Since both the MPPs of interest are strongly influenced by cohort effects, we assume different baseline hazard functions, specific for birth cohort, C: the cohorts considered are 1946–1955, 1956–1965, 1966–1975. This construction avoids the assumption of proportional hazards for the variable cohort, which is in fact unrealistic, supposing a different baseline behavior for each cohort.

The explanatory variables chosen to describe the original family characteristics are: Z_1 indicating the father's education level (1 = Secondary or higher, 0 = otherwise) and Z_2 for mother's working experience (1 = always or nearly always employed, 0 = otherwise).

To construct the chain graph, the set of nodes is $\{X(t, m_X), Y(t, m_Y), U, Z_1, Z_2, C\}$: variables and processes are partitioned into an ordered sequence of three blocks, with Z_1, Z_2 and C in the block on the right representing pure explanatory variables; $X(t, m_X)$ and $Y(t, m_Y)$ in the block on the left representing the pure responses, whereas the unobserved variable U is in the intermediate block.

Figure 1 shows the posterior distributions for γ coefficients; these do not give evidence in favor of zero. Consequently $X(t, m_X) \not\perp\!\!\!\perp_{\rightarrow} Y(t, m_Y) \mid U, Z_1, Z_2, C$ and

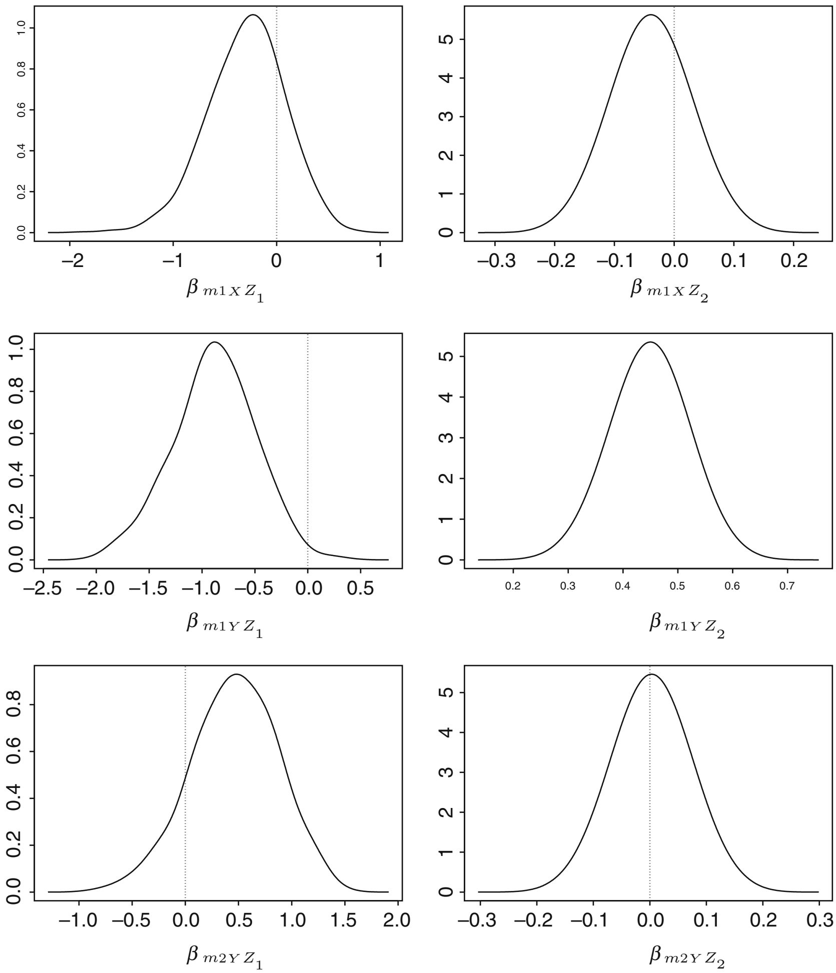

Fig. 2 Posterior distributions of β regression parameters

$Y(t, m_Y) \not\!\perp\!\!\!\perp\!\!\!\rightarrow X(t, m_X) \mid U, Z_1, Z_2, C$ and the nodes of MPPs are connected by an undirected edge, given that the two MPPs are interdependent.

Figure 2 illustrates the posterior distributions for β regression parameters: $\beta_{m1_XZ_1}$ and $\beta_{m1_XZ_2}$ the coefficients for the MPP "to have a child", $\beta_{m1_YZ_1}$ and $\beta_{m1_YZ_2}$ the coefficients for the MPP "to find a job" and $\beta_{m2_YZ_1}$ and $\beta_{m2_YZ_2}$ for the MPP "to leave a job" respectively for the covariates Z_1 "father's education" and Z_2 "mother's working experience". Results suggest that "father's education" and "mother's working experience" are strongly related "to find a job", but not "to leave a job" and "to have a child". By definition (1), this implies that $X(t, m_X) \perp\!\!\!\perp (Z_1, Z_2) \mid Y(t, m_Y), U, C$, but $Y(t, m_Y) \not\!\perp\!\!\!\perp (Z_1, Z_2) \mid X(t, m_X), U, C$.

Figure 3 shows posterior distributions of parameters U representing individual (one for each woman) frailty coefficients, for each mark. The dissimilarities on the posterior distributions stress the different frailties of women.

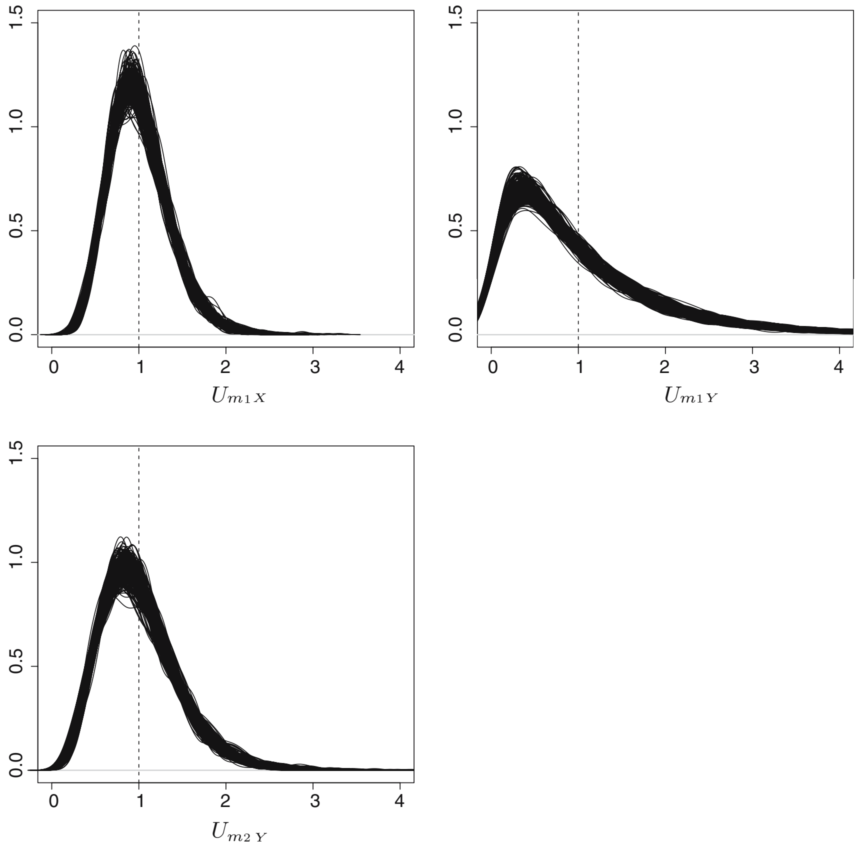

Fig. 3 Posterior distributions of parameters U representing individual frailty for each mark

Figure 4 describes the birth cohort specific baseline hazard functions for the fertility process. The birth cohort specific baseline hazard functions for female labor force participation processes, not reported here, present two peaks for "to find a job" process: respectively for 14 and 18-years and 20 and 26-years on the first and second cohort. The process for "to leave a job" presents a peak at 18-years for the cohort 1946–1955 and at 24-years for the cohort 1956–1965, reflecting conventional educational courses end and common precariousness of first jobs. Results about the third cohort are less relevant because of censoring. The non-parametric part of the mark-specific hazard functions are therefore able to capture these particular moments of women life. Note that, in case of the process being independent of the cohort, the three baseline hazard functions are undistinguishable. The heavy differences among them suggest a strong dependence of the fertility process on the cohort. Same results have been yielded for the labor-force participation process.

Figure 5 illustrates the resulting graph, where, for shortness and legibility, are omitted the prior parameters and hyperparameters nodes. The missing arrows

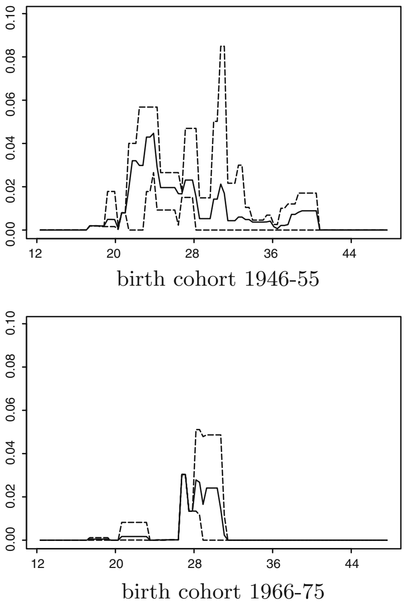

Fig. 4 The estimates of the birth cohort specific baseline hazard functions with their 95% credibility interval, for the fertility process

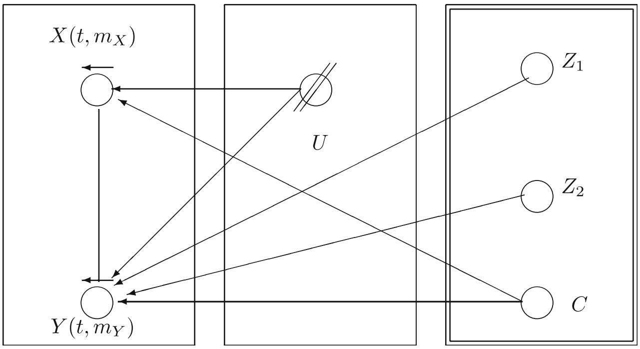

Fig. 5 The resulting graphical duration model

from Z_1 and Z_2 to $X(t, m_X)$ indicate the conditional independencies mentioned above. Note that, for assumption, the latent variable U is independent of the other explanatory variables.

These results are consistent with the previous analysis (in Gottard 2006) with respect to the presence of interdependence between the two processes, even if the results are not fully comparable. As a matter of fact, the entire data set and not only one cohort has been considered here. Moreover, model assumptions, such as covariate included and functional form of the hazards, are completely different. With respect to the model in Gottard (2006), the semiparametric specification adopted here, does not require strong assumptions on the form of the hazard and on the specification of the relationship of processes on the cohort effect. Furthermore, this Bayesian model allows for the inclusion of the frailty terms in a very easy way.

5 Discussion and conclusions

This paper proposed a first attempt to study the interdependence between two MPPs using a Bayesian graphical duration model. Bayesian approach is a very flexible and useful instrument, allowing for a semi-parametric implementation of the mark-specific hazard functions and individual frailty. This is particularly appealing for studying dynamic systems with a complex association structure and a lot of potential influencing variables and processes. The model presented allows to avoid the hypothesis of proportional hazard for some explanatory variables and assume it for others, giving a great flexibility. Graphical duration models has confirmed their usefulness in illustrating the relationship among MPPs and random variables in a precise and direct way.

Acknowledgements We are grateful to INF-2 group for having kindly made available the data set used in the present work. This work was presented at S.Co.2003 - *Modelli Complessi e Metodi Computazionali Intensivi per la Stima e la Previsione*, in Treviso, Italy, September 2003. The authors would like to thank Alberto Roverato and the S.Co.2003 participants for their constructive comments and suggestions.

References

Aalen OO (1987) Dynamic modeling and causality. Scand Actuar J 12:177–190
Arjas E (1989) Survival models and martingale dynamics (with discussion). Scand J Stat 16:177–225
Arjas E, Gasbarra D (1994) Nonparametric Bayesian inference from right censored survival data, using the Gibbs sampler. Stat Sin 4:505–524
Blossfeld HP, Klijzing E, Pohl K, Röhwer G (1995) Modelling parallel processes in demography. An application example of the causal approach to interdependent systems. In: 3rd European population conference on event history analysis in demography, Milano
Cox DR, Lewis PAW (1972) Multivariate point processes. In: Proceedings of the Berkely symposium on mathematical statistics and probability, vol III.
Cox DR, Wermuth N (1996) Multivariate dependencies. Models, analysis and interpretation, Chapman & Hall, London
Dawid AP (1979) Conditional independence in statistical theory (with discussion). J R Stat Soc B 41:1–31
De Sandre P, Ongaro F, Rettaroli R, Salvini S (1997) Matrimonio e figli: tra rinvio e rinuncia, Il Mulino, Bologna
Frydenberg M (1990) The chain graph Markov property. Scand J Stat 17:333–353
Gottard A (2006) On the inclusion of marked point processes in graphical models. A draft can be found in http://www.ds.unifi.it/ricerca/pubblicazioni/working_papers/2002/wp2002_07.pdf (submitted)
Härkänen T (2002) BITE: Bayesian intensity estimator reference manual. (Online user manual, http://www.rni.helsinki.fi/~tth/bite.html)
Härkänen T, Virtamen JI, Arjas E (2000) Caries on permanent teeth: a non-parametric Bayesian analysis. Scand J Stat 27(4):577–588
Ibrahim JG, Chen M-H, Sinha D (2001) Bayesian survival analysis, Springer-Verlag, Berlin Heidelberg New York
Lauritzen SL (1996) Graphical models. Oxford Science Publications, Oxford

Lauritzen SL, Wermuth N (1989) Graphical models for the association between variables, some of which are qualitative and some quantitative. Ann Stat 17:31–57

Moro A, Gottard A (1999) Analisi di biografie mediante processi di punto: approccio “sistemico” e “causale” a confronto. In: De Sandre P, Pinnelli A, Santini A (eds), Nuzialità e fecondità: i percorsi del cambiamento. Il Mulino, Bologna

Schweder T (1970) Composable markov processes. J Appl Probab 7:400–410

Tuma NB, Hannan MT (1984) Social dynamics, models and methods. Academic, Orlando